LIRIS

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

FRE 2672

# Les ontologies et leurs méthodes de construction

Catherine ROUSSEY

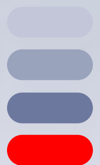
UCBL

VISIT…

LANZAROTE Caliente.COM

# Ontologie: Objectifs

- Systèmes experts → systèmes à base de connaissances
- Séparer les connaissances, des traitements sur ces connaissances pour résoudre un problème, effectuer une tâche
- Réutilisation des connaissances dans différents systèmes



# Ontologie: Définitions

- **Gruber 1993: «** *Une ontologie est une spécification explicite d'une conceptualisation d'un domaine.* **»**

- **Studer 1998 : «** *Une ontologie est une spécification formelle et explicite d'une conceptualisation partagée***.** **»**

  - **conceptualisation =** Modèle abstrait de phénomènes réels,
    Ensemble des concepts importants caractérisant un domaine,
    Point de vue, une perception limitée du monde
  - **explicite :** les types de concepts utilisés et les contraintes sur leurs usages sont explicitement définis**.**
  - **formel :** compréhensibles par la machine.
  - **partagé :** connaissances consensuelles, acceptés par un groupe

LIRIS

# Ontologie : Composants

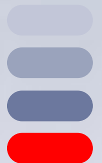
- **Des termes:** un vocabulaire normalisé
- **Des concepts/classes/types de concepts:** donne le sens de termes par des définitions en langages naturelles.
- **Des relations entre concepts**
- **Une taxonomie de concepts**
- **Des propriétés/attributs de concepts**
- **Des définitions formelles** dans un langage de représentations des connaissances (Frames / LD …)
- **Des axiomes/contraintes d'intégrités**
- **Des fonctions** (pour calculer une valeur)
- **Des règles** (si X alors Y)
- **Des instances/individus**

**Lightweight Ontology**

**LO**

**Heavyweight Ontology**

**HO**



# Ontologie : Langages de représentation

- Réseaux sémantiques
- Graphes
- RDF

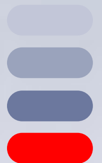
Lightweight Ontology

LO

- Logiques de Description
  LOOM
- Langage à base de Frames + logique du 1er ordre
  KIF/OCML/FLogic
- Langage de balise (HTML/XML)
  SHOE/RDFS/OIL/DAML/OWL

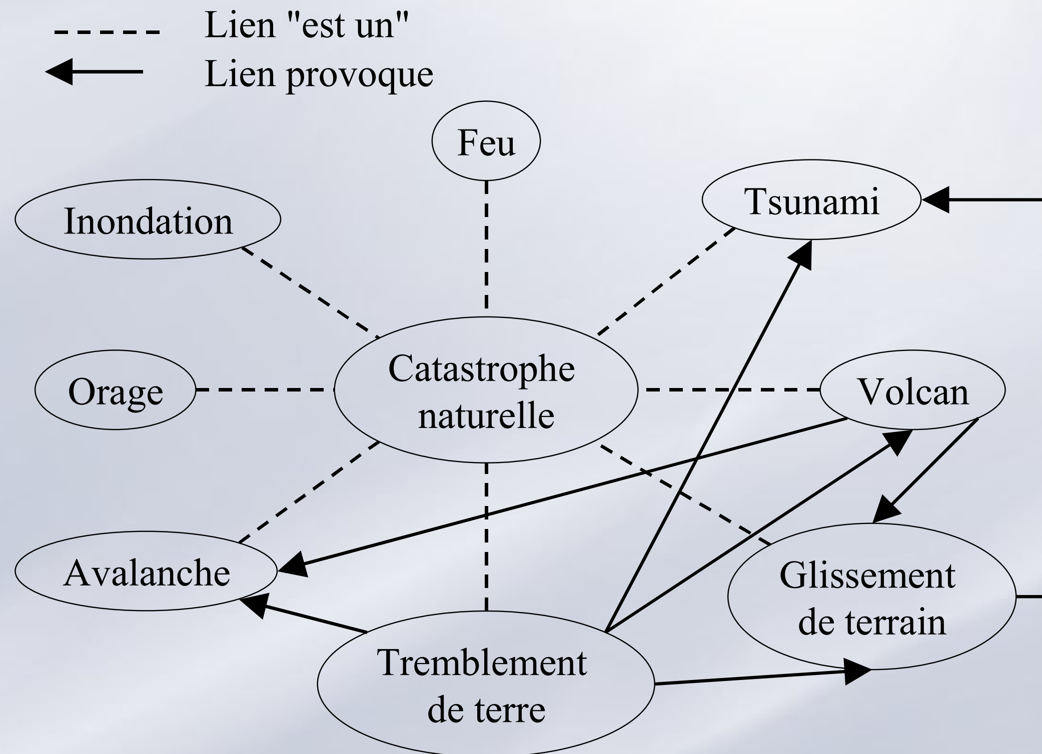
Heavyweight Ontology

HO



# Ontologie : Exemples (LO)

- Wordnet
- Eurowordnet
- Sensus
- Mikrokosmos
- Thésaurus
- UMLS (thésaurus médical)
- Yahoo Directory



# Ontologie: Usage (LO)

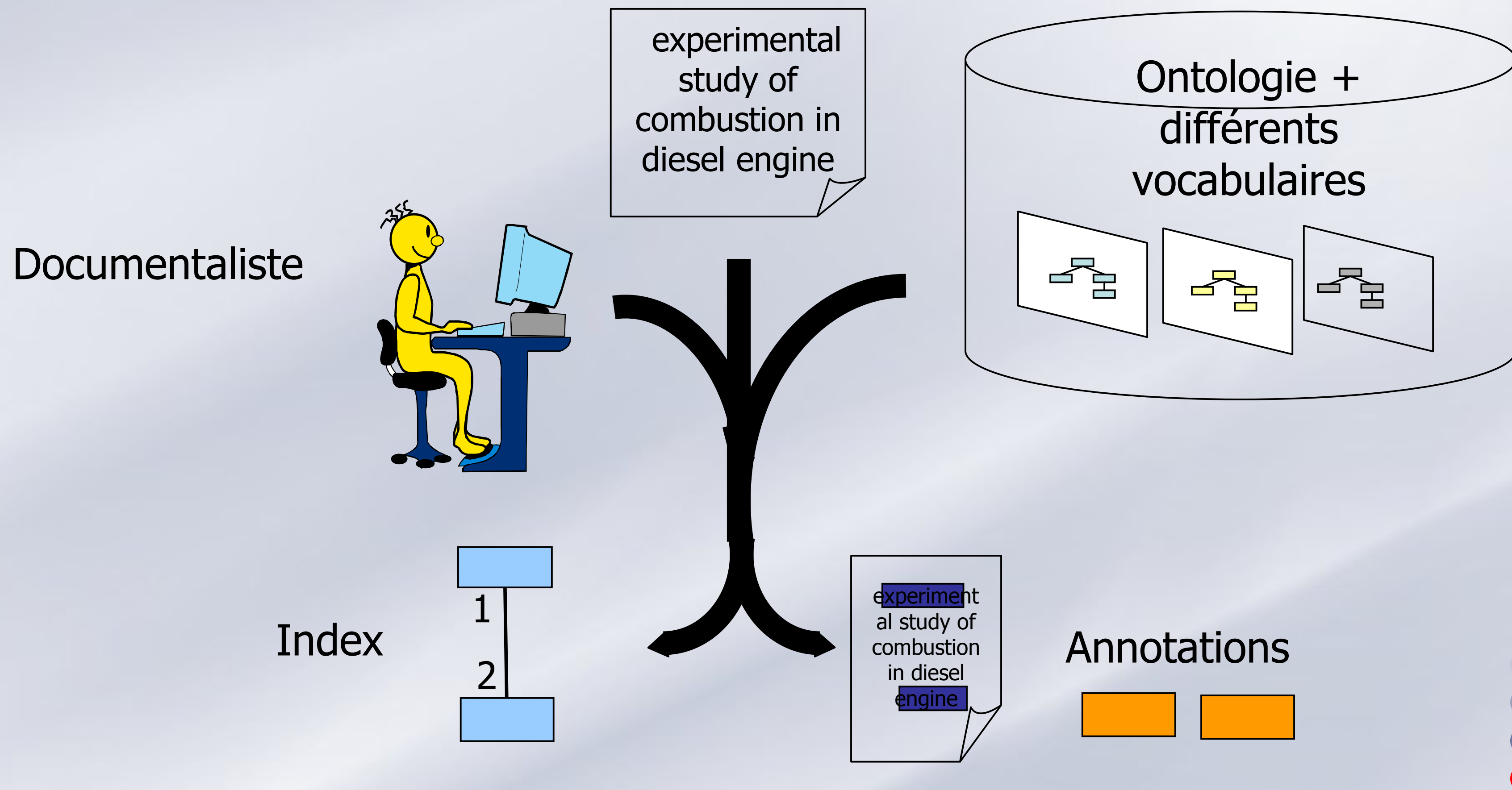
- **Système de Traitement Automatique du Langage Naturel (TALN)**
  **ontologie linguistique**
- **Système de traduction**
  **ontologie multilingue**
- **Systèmes de Recherche d'Information (RI)**
  **thésaurus** -----

**Exemple en RI**

- **Phase d'indexation:**
  Représenter le contenu du document avec les termes de l'ontologie
- **Phase de recherche:**
  Calculer la similarité entre les termes des documents et les termes de la requête à l'aide des liens de l'ontologie.

# Ontologie: Logiciel SyDoM

SyDoM Système Documentaire Multilingue: Roussey 2001



# Ontologie: Logiciel SyDoM

Recherche multilingue de documents XML

Fichier Concept Relation Recherche

francais

Arbre XML GC arbre GC XML GC

création
bruit
1
2
1
resultat
origine
2
moteur diesel

Concepts

modélisation
évaluation
processus physique
bruit
combustion
création

Relations

patient
thème
source
initiateur
agent
origine
ressource
pièce
étape

Recherche des documents correspondant à la requête

C:\testjava\DocsXML\noise18.xml : 1.0
C:\testjava\DocsXML\noise19.xml : 0.5
C:\testjava\DocsXML\slap20.xml : 0.5
C:\testjava\DocsXML\acous32.xml : 0.5
C:\testjava\DocsXML\acous33.xml : 0.5

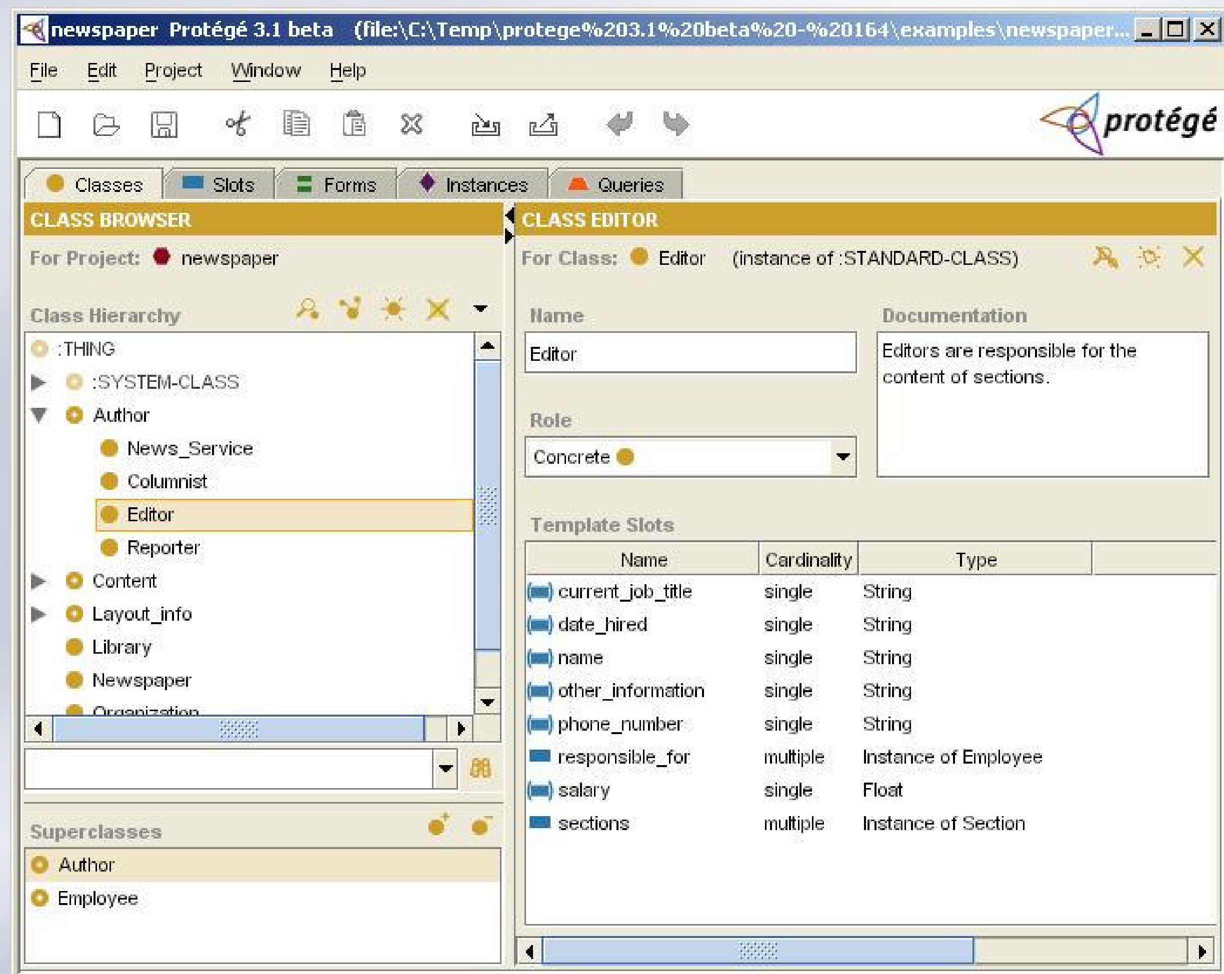
Recherche des documents correspondant à la requête

| Nom | Valeur |
|---|---|
| C:\testjava\DocsXML\noise18.xml | 1.0 |
| C:\testjava\DocsXML\noise19.xml | 0.5 |
| C:\testjava\DocsXML\slap20.xml | 0.5 |
| C:\testjava\DocsXML\acous32.xml | 0.5 |
| C:\testjava\DocsXML\acous33.xml | 0.5 |

Mise a jour du fichier inverse : fin

# Ontologie : Exemples (HO)

- Cyc ontology
- KA² initiative
- GALEN : Ontologie médicale
- Enterprise Ontology



# Ontologie: Usage (HO)

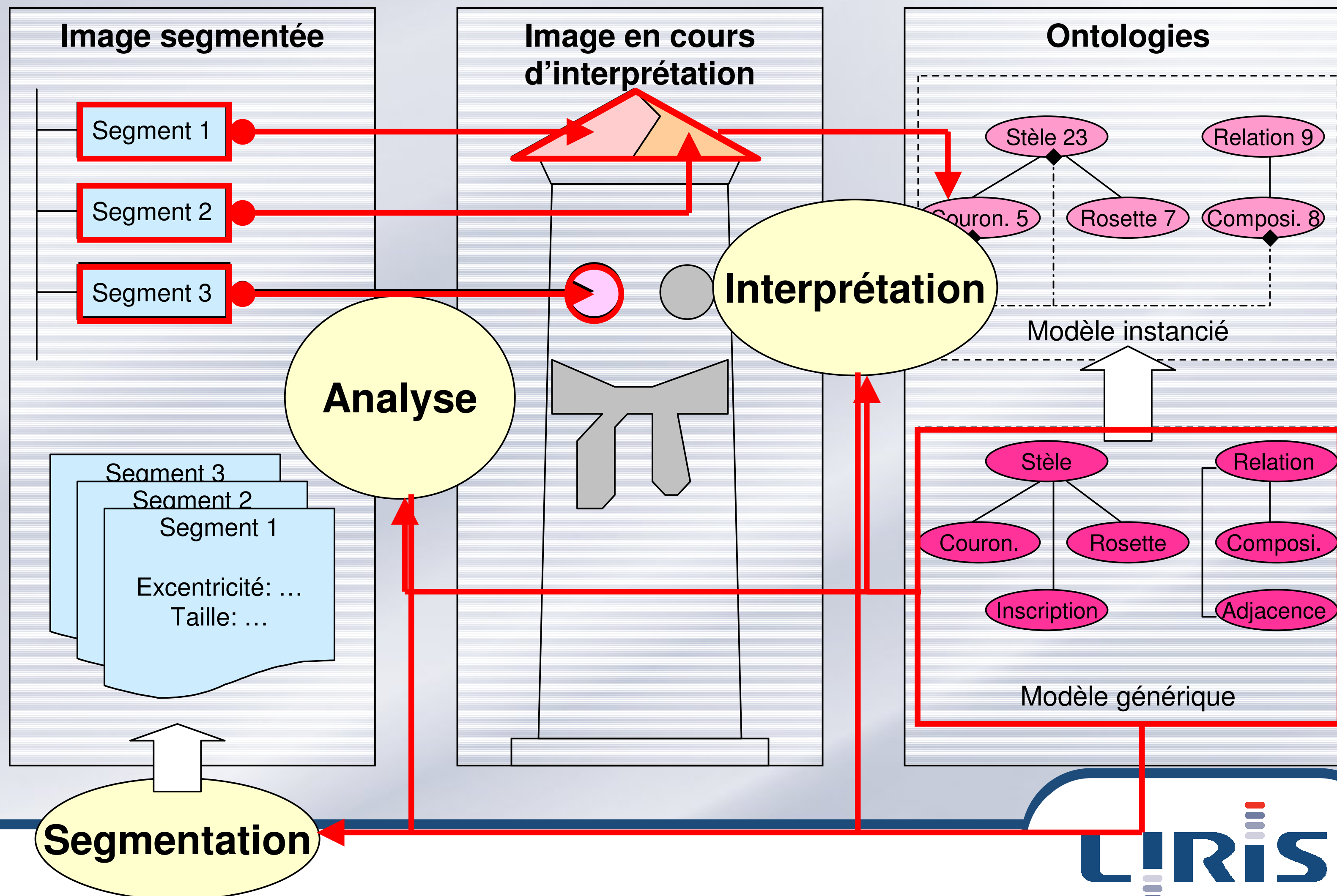
- Inférer de nouvelles connaissances/données
- Système à base de connaissance dédié à la résolution d'un problème
  - Interprétation d'image
  - Aide au diagnostic d'un malade/ d'un problème technique/ d'un exercice de mathématique…
- Interopérabilité des Systèmes d'Informations
  - Mise à jour de données
  - Découverte d'incohérences
  - Médiation / collaboration
- Knowledge Management / Web sémantique

Exemple en Interprétation d'image

- Diriger la segmentation d'une image pour découvrir la sémantique des objets issus de la segmentation

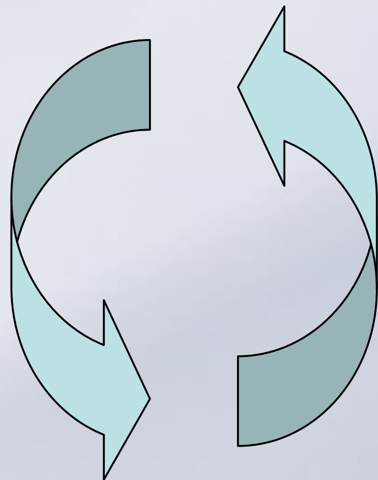
# Ontologie: Interprétation d’Image

Zlatoff, Tellez, Baskurt 2003

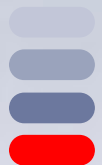


# Méthodes de construction d'ontologies

## Cycle de vie en plusieurs étapes



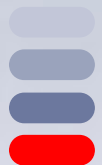
- Spécification
- Acquisition de Connaissances
- Formalisation
- Évaluation
- Documentation



# Spécification

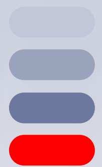
a quoi va servir l'ontologie?

- Identifier le but de l'application
- limiter  le domaine
  - les objets
  - les usagers, les points de vues
  - les sources documentaires
- Trouver les questions auxquelles devra répondre l'ontologie.
- Trouver les scénarios d'utilisation des connaissances



# Acquisition des connaissances

- Réutiliser des ontologies existantes
  - top level ontology
  - Ontologie d'une partie du domaine ex. ontologie des unités
- Identifier les termes importants
  - normaliser le vocabulaire
- Identifier les concepts et les relations du domaine
  - Définition écrite en Langue Naturelle
  - Trouver les conditions minimales et suffisantes pour dire qu'un objet appartient à une classe donnée.



# Construire la taxonomie de concepts

- Hiérarchie avec une relation is a ou relation de subsomption
- 3 stratégies pour identifier les concepts
  - Bottom-up : spécialisé → général
    concepts très détaillés pas nécessairement utiles
  - Top-down: général → spécialisé
    facile la cohérence et la réutilisation mais dirige la conception suivant un point de vue
  - Middle-out: concept important →spécialisé, général
    ontologie modulaire, facilite la stabilité des résultats

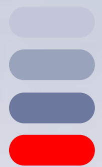
# Formalisation

- Coder l'ontologie dans un langage formel
- des outils: Protégé, Kaon, OntoEdit …
- Trouver les classes, les attributs, les types, les contraintes
- Peupler l'ontologie: instancier les classes



# Evaluation

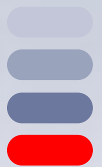
- Construction d'ontologie: un domaine récent
- Valider la taxonomie:
  - Pas de cycle
  - Toutes les instances d'une classe sont aussi les instances de la classe mère
  - Hiérarchie homogène: pas de classe isolée, pas de listes de sous classes trop importante
- Tester l'application



LIRIS

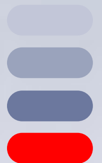
# Documentation

- Donner des explications
- Définition en langage naturel
- Lier les concepts aux sources dont sont issues les définitions



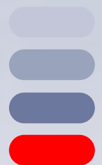
# Conclusion

- Toutes les méthodes sont faites pour des experts en ontologies qui communiquent avec des experts du domaine
- La phase d'acquisition des connaissances est la plus longue et la plus fastidieuse.
- Écrire une ontologie valide n'est pas évident
  - Différence entre un concept et un terme, une instance et une classe, …
  - Trouver un consensus entre experts du domaine
  - La relation partie-de n'est pas la relation hiérarchique de la taxonomie



# Bibliographie


- Gómez-Pérez, A. (1998). Knowledge sharing and reuse. Handbook of Applied Expert Systems. Liebowitz, editor, CRC Press.
- Gómez-Pérez, A., Fernandez-Lopéz, M., Corcho, 0 (2003). Ontological Engineering, Springer Editor.
- Uschold, M. and Gruninger, M. (1996). Ontologies: Principles, Methods and Applications. Knowledge Engineering Review 11(2)
- Gruber, T. R. Toward principles for the design of ontologies used for knowledge sharing. Padova, Italy, 1992.

LIRIS

LIRIS

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

FRE 2672

EDU

# Le projet Towntology


LIRIS: Kader K., Laurini R., Roussey C.

EDU: Berdier C.,Toussaint JY., Zimmerman M.

Développeurs: Beaulieu C., Berthet C., Tardy Y., Gesche S.


UCBL

INSA LYON

# Objectifs

- Construction d'une Ontologie:
  - Du domaine de l'urbanisme
  - En français
  - Pour l'interopérabilité des BD, des SIG

- Réalisation de plusieurs maquettes
  - Projet de 5 étudiants EDU.
  - Plus de 1000 termes
  - 21 puis 12 types de relations identifiées entre les termes
  - Aide à la clarification du vocabulaire
  - Construction d’un outil de visualisation d’une ontologie non formelle en cours de construction

    → plusieurs définitions par concepts

# Démarche suivie

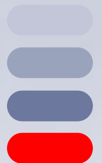
- Acquisition des connaissances
  - Trouver les ressources documentaires adéquates
  - Interroger des utilisateurs potentiels
- Sélectionner les termes importants du domaine.
- Définir ces termes: textes et photos
- Construire un vocabulaire
- Trouver les concepts structurants !
- Définir les relations nécessaires pour organiser les concepts entre eux

  valider ces relations

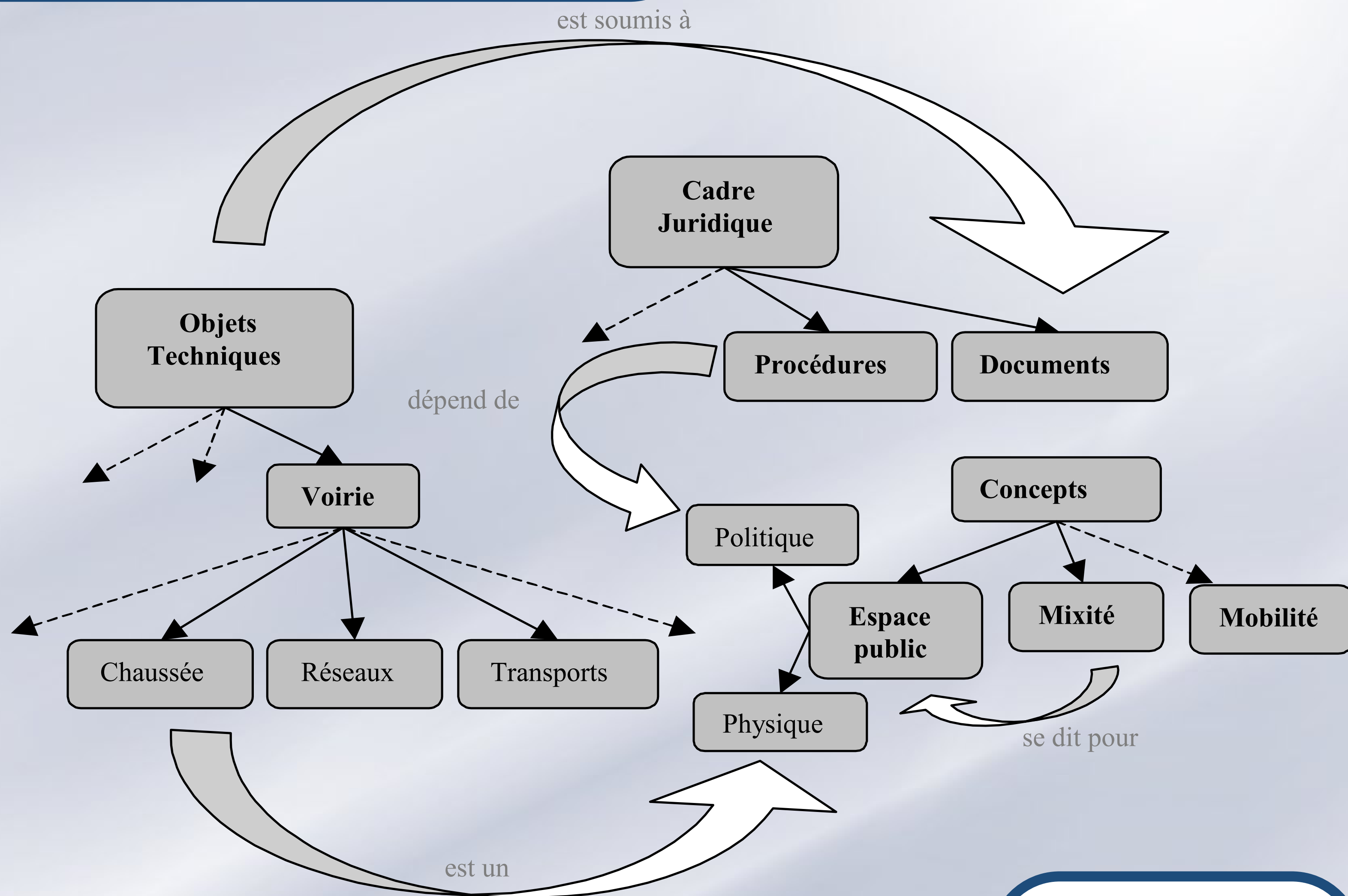
# Les ressources

- Les ressources utilisées
  - Dictionnaire de l’urbanisme et de l’aménagement
  - Grand dictionnaire terminologique
  - Le web
  - Des questionnaires

- Les ressources disponibles
  - Des contacts professionnels
  - Des schémas de BD



# Les domaines abordés



# Les définitions

**SABLE**

(Généralité) Au sens courant, matériau meuble formé de grains de quartz (grains de sable) tel celui des plages ou des dunes.

(Géotechnique) Ensemble de grains ou de menus fragments de minéraux ou de roches, d'un diamètre compris entre 0,02 et 2 mm. On distingue le « sable fin » (fine sand) constitué de particules d'un diamètre compris entre 0,02 et 0,2 mm et le « sable grossier » (coarse sand) constitué de particules d'un diamètre compris entre 0,2 et 2 mm.

Office de la langue française



Source : http://www.fmg-hydro.com/photos/sable.jpg

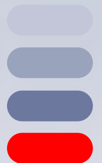
*En savoir plus*

***Figure 1 : exemple de définition du terme* sable.**

LIRIS

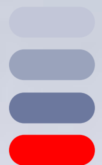
# 21 types de relations

- A pour matériaux
- Est composé de
- Est un, Est un sous-ensemble de
- Est un outil pour, Est un outil (matériel) (étude) (informatique) pour.
- S'oppose à
- Est utilisé pour
- Est situé sur, Est situé dans
- Est une procédure concernant , Est une opération pour, Est un document pour
- A une activité concernant
- Travaille pour
- Caractérise, Dépend de, Est issu de
- Est un problème touchant
- Se dit pour



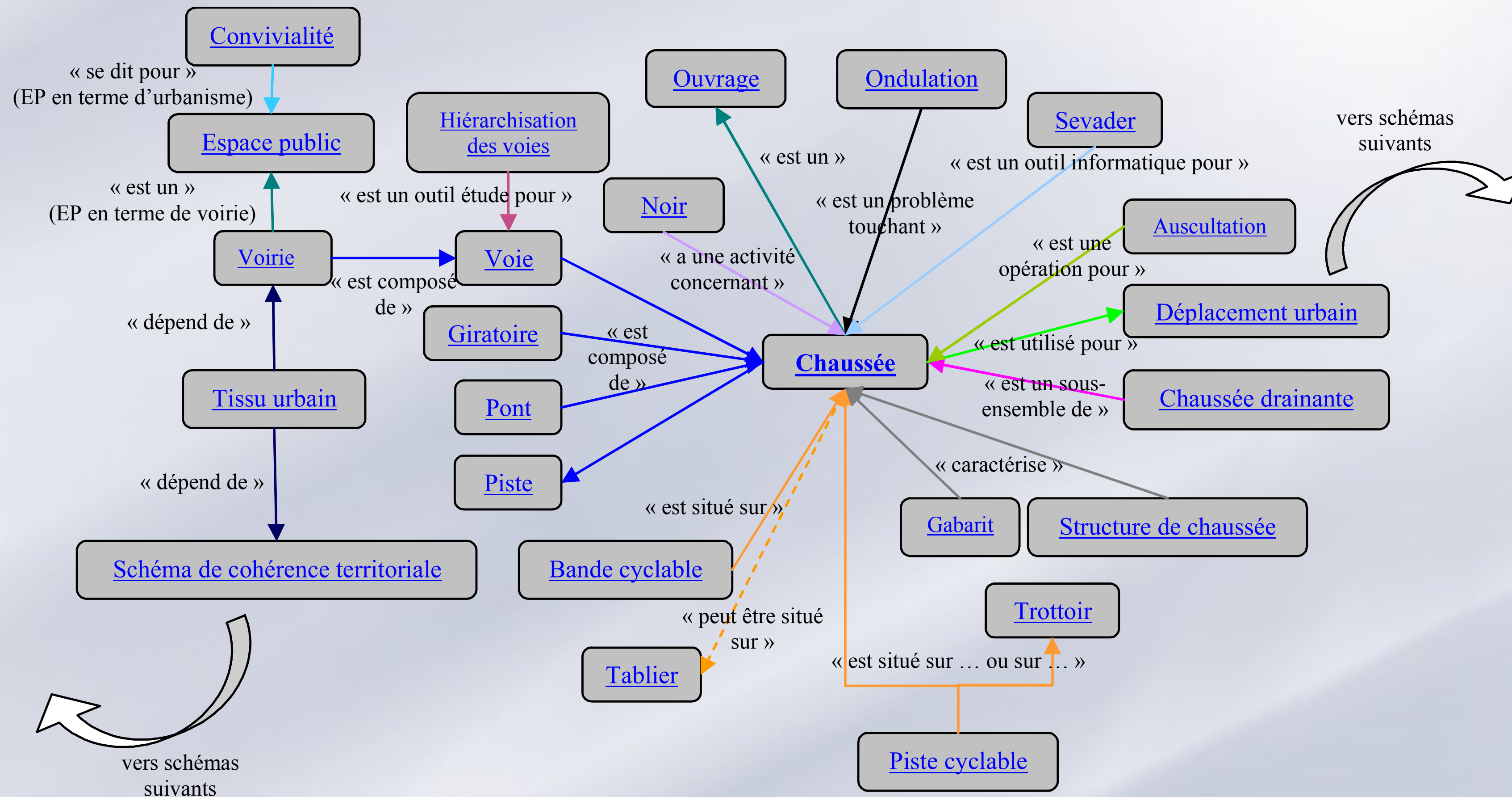
# 12 types de relations

- *Relation de subsomption ou d'héritage:* Est un
- *Relations de composition :* Est composé de
- *Relation de typage et de structuration:*
  - Se dit pour
  - Est caractérisé par
- *Relations topologiques*
  - Est situé sur
  - Est situé dans
- *Relation de causalité:* Est issu de
- *Relation de subordination:*
  - Dépend de
  - Travaille pour
- *Relations d'utilisation*
  - Tient le rôle de
  - Est utilisé pour
  - Est utilisé par

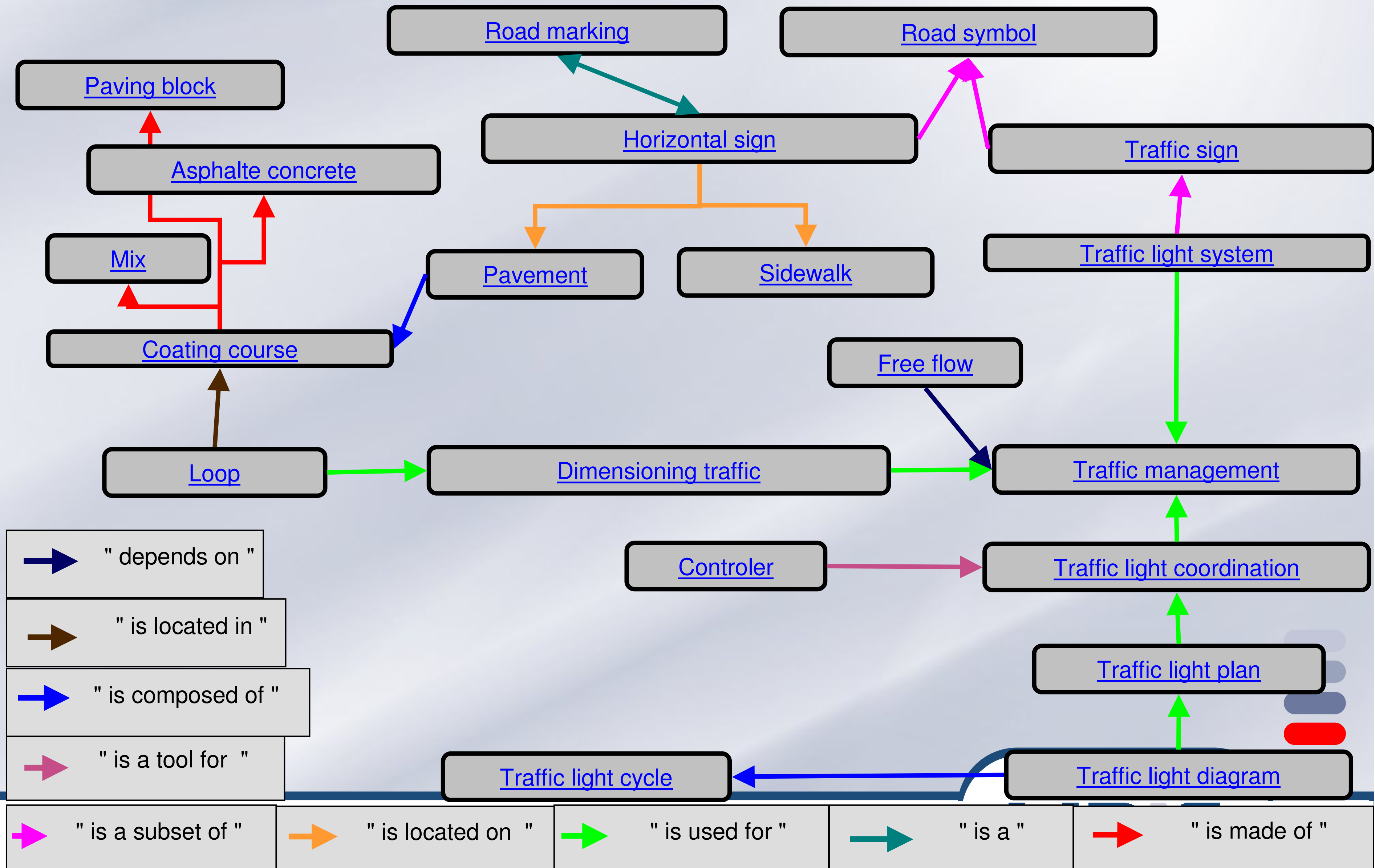


LIRIS

# Le réseau sémantique



# Exemple d'une partie de l'ontologie



# Conclusion

- Acquisition des connaissances par les experts du domaine et non par les ontologistes
- Problème
  - Manque de structuration (ontologie de haut niveau)
  - Limiter le nombre de relations
  - la formalisation ?

- Construction d'un hypertexte sémantique du domaine
- Définir une DTD XML valide pendant la construction
- Projet européen COST (ontologie multilingue)
  13, 14 juin à Lyon

LIRIS